DOCUMENT RESUME

ED 250 691 CS 208 647

| | |
|---|---|
| AUTHOR | O'Donnell, Holly |
| TITLE | Large Scale Writing Assessment. ERIC Digest. |
| INSTITUTION | ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill. |
| SPONS AGENCY | National Inst. of Education (ED), Washington, DC. |
| PUB DATE | 84 |
| CONTRACT | 400-83-0025 |
| NOTE | 14p.; Provided in both typewritten version and one-page typeset version. |
| PUB TYPE | Information Analyses - ERIC Information Analysis Products (071) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Educational Assessment; Educational Trends; Elementary Secondary Education; Evaluation Methods; *Holistic Evaluation; *Scoring; *Test Reliability; Writing (Composition); *Writing Evaluation |
| IDENTIFIERS | *Analytical Scoring; ERIC Digests; *Primary Trait Scoring; Theory Practice Relationship |

ABSTRACT

Intended for administrators and policymakers as well as teachers, this digest explores approaches to, problems with, and trends in large scale writing assessment. The digest first compares direct and indirect assessment approaches, then examines three approaches to scoring: holistic, primary trait, and analytic. It then discusses issues and problems surrounding the reliability and validity of the scores generated by such assessment. Finally, the digest examines trends characterizing large scale writing assessment. (ETH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

Holly O'Donnell

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED250691

ERIC Digest

# Large Scale Writing Assessment

A national concern over the decline in students' writing scores, (as revealed in National Assessment of Educational Progress reports), serious doubts about what some of those scores signify, and a shift in focus from writing product to writing process in research and classroom practice have each given impetus to the change from indirect measures of writing proficiency (those that use "objective" test items) to direct measures (those that call for student writing samples). In their 1981 national survey, McCready and Melton found that of the 24 states claiming to have a writing assessment program, 22 require a writing sample as part of the assessment. Only two states rely solely on the use of objective tests.

Large scale writing assessments, however, involve a number of complex issues that are not always evident to decision-makers who are not specialists in measurement. In discussing how to and how not to conduct an assessment of student writing, McCaig (1982) warns that "an assessment plan which is incomplete or poorly conceived may produce findings which can be challenged and even dismissed as meaningless by critics who can document flaws in the process." This digest (1) outlines some of the approaches used in the implementation of large scale writing assessments, (2) examines some of the issues and problems surrounding the use of student writing samples, and (3) reports on various trends in state writing assessment projects.

## DIRECT VERSUS INDIRECT ASSESSMENT

Direct and indirect writing assessments are radically different approaches focusing on different components of writing. Indirect

CS208647

measurements typically use multiple choice tests to assess the student's understanding of mechanics or language conventions: spelling, punctuation, capitalization, grammar, usage, sentence construction, organization, and so on. Direct assessments, on the other hand, assess actual writing performance, since they require the students to produce a writing sample. Spandel and Stiggins (1981) suggest that the two approaches can best be compared in terms of their advantages and disadvantages.

| | Advantages | Disadvantages |
|---|---|---|
| Direct | • extent of information provided about the students' writing proficiency | • potential lack of uniformity regarding proficiencies assessed |
| | • fidelity to real world writing tasks | • high cost of scoring |
| | • potential for positive user attitudes | |
| | • relatively low test development cost | |
| | • high face validity | |

| | | |
|---|---|---|
| Indirect | o high score reliability | o lack of fidelity to real world writing tasks |
| | o relatively low scoring cost | o reliance on reading |
| | o high degree of control over skills tested | o lack of face validity |

Participants at a conference on assessment issues agreed that the use of writing samples is essential because of the instructional implications(McCready and Melton, 1981). That is, if teachers know that students' writing ability will be evaluated by means of a direct measure, they will encourage more writing in the classroom.

THREE APPROACHES TO SCORING

Provided that writing assessments are conducted to determine the status of writing in a given state or school system and provided they are conducted to help improve classroom instruction, several factors need to be considered prior to the collection of writing samples: (1) the educational decisions to be made, on the basis of test results; (2) the writing purpose, audience, and type of writing to be required; and (3) the specific skills or traits to be judged along with the criteria used for evaluating writing performance (Spandel and Stiggins, 1981). It must also be remembered that ratings will vary depending upon the scoring procedure used. Quellmalz, in writing about scoring criteria (1982), notes that "criteria employed for evaluating student writing vary along a number of dimensions: from qualitative to quantative; from general to specific; from comprehensive, full discourse features to isolated features; from vague guidelines to

discourse features to isolated features; from vague guidelines to replicable, objective guidelines." Scoring options range from holistic scoring (general impressionistic marking) to analytic and primary trait scoring.

## HOLISTIC SCORING

Holistic scoring of a writing sample is based upon the reader's overall impression of the effectiveness of a piece of writing. Papers are scored by trained raters on a numerical scale, usually a four-point scale. Once the writing samples are collected, the raters or scorers sort the samples into four stacks, relating the quality of the essay only to other papers in the group rather than to a predetermined example of "good" writing. Papers are typically read by two raters and the scores they assign a writing sample are summed into a total score. If there is a discrepancy of two score points, the score is reconciled by yet a third reader/rater.

## PRIMARY TRAIT SCORING

Primary trait scoring focuses on a specific rhetorical characteristic or trait of a given piece of writing. It is based on the premises that all writing is done in terms of a specific audience and that writing, if successful, will have the desired effect on that audience. Lloyd-Jones (1977) identifies two goals of primary trait scoring: (1) to define what segment of discourse will be evaluated (e.g., explanatory, expressive, persuasive), and (2) to train readers to render holistic judgments accordingly. A scoring guide for primary trait analysis may consist of the exercise itself; a description of the rhetorical traits of the writing; an interpretation of the exercise indicating how each element in the task is expected to affect the student; an interpretation of how the situation of the exercise is related to the primary trait; sample papers that are

representative of each score point; and a discussion of why each sample paper was scored as it was (McCready and Melton, 1981). One difference between holistic and primary trait scoring is that with primary trait scoring, students' papers are being measured against external criteria, whereas with holistic scoring, papers are compared with one another.

## ANALYTICAL SCORING

If primary trait scoring is a situation-specific analysis of writing, analytical scoring is a thorough, trait-by-trait analysis. The identified traits reflect those components of a writing sample that are considered important to any piece of writing in any context. Diederick (1974), the originator of analytical scoring, for example, has identified eight common traits: ideas, organization, wording, flavor (tone), usage, punctuation, spelling, and handwriting. Others may use traits more general such as content, organization, focus and support, and mechanics. If enough components are analyzed, this scoring procedure can provide a comprehensive picture of writing performance. However, the components need to be explicit and well defined so that the raters understand and agree upon the basis for making judgments about the writing sample.

In relating these scoring approaches to classroom applications, Spandel (1981) observes that holistic scoring offers a broad base for a discussion of what makes a piece of writing generally good or bad. Analytical scoring can take this discussion one step further by identifying those traits of components that make a piece of writing effective. And, by being situation-specific, primary trait scoring focuses on the importance of audience to a piece of writing.

## ISSUES AND PROBLEMS

Essential to the quality of assessment and the value of scoring procedures used are the reliability and validity of the scores generated by the assessment. Specifically, the scoring criteria should be applicable uniformly within a rating session and from one rating session to another. Furthermore, these ratings should correlate with other measures of student writing. Even if the assessment instrument is reliable and valid, spurious scores can result from the development of poor exercises, poor test administration or environment, or poor scoring procedures (Stiggins, Northwest Regional Educational Laboratory, 1980). Scorers must understand and agree upon the criteria applied to a piece of writing. Thus, inadequate training of scorers may also influence or skew the results of an assessment.

The choice of topic (or "prompt") to be written is another factor that may influence scores. Students may write more enthusiastically on some topics than on others, resulting in better quality writing. A student's background and prior knowledge will also affect the "expertise" a student brings to a piece of writing. Or, depending upon the student's interpretation of a writing prompt, he or she may write persuasive discourse in response to a prompt intended for expository discourse.

Time and cost are two other factors that may influence the decision for large scale writing assessment. Stiggins (Northwest Regional Educational Laboratory, 1980) separates cost factors into those that are developmental and those that are administrative. Developmental costs will vary depending on whether a previously designed assessment instrument is used or a new one developed. If one is to forego the expense of constructing a new scoring instrument, expense will still be incurred for the securing, reviewing, evaluating, and selecting of appropriate exercises and scoring guides that do exist. Administrative costs involve those associated with test

collected by Quellmalz (1982), the training time for holistic and primary trait scoring averages two to four hours, and for analytical scoring averages six to eight hours. Test reuse is another cost factor. Stiggins (Northwest Regional Educational Laboratory, 1980) states that "with indirect assessment, the high cost of test construction can be amortized over repeated administration and the recurring scoring costs are very low. However, with direct assessment, although the initial development costs are low, the scoring costs remain high with each use." In a 1982 dissertation Bauer compared the reliabilities and the cost-efficiencies of these three methods of direct assessment. Based on her results, Bauer concluded that the analytical method was the most reliable and the holistic method was the most cost-efficient in grading large number of essays (see ED 225 171).

A 1979 study by Fredrick identified some of the problems that states have encountered with their writing assessment endeavors: (1) arrangements for a place large enough and suitable enough for the scoring, (2) coordinating release time for teachers to act as scorers, (3) adhering to a rigid time schedule during the scoring session, (4) not enough time or money, and (5) finding or designing a variety of writing skill tasks. Her survey concluded with a list of recommendations to others who are planning a writing assessment, some of which are as follows:

- formulate writing objectives and focus research questions before the assessment,
- use language arts specialists to advise on content and to react to items prior to pilot testing,
- clarify traits to be measured,
- include clear and concise directions, and
- use actual performance on practical writing, such as as messages, letters, forms and so forth, instead of the proofreading type of assessment found in most multiple choice tests.

TRENDS IN WRITING ASSESSMENT

A national study conducted in 1981 by McCready and Melton collected data from 42 state departments of education. Of the 22 states using a writing sample, most of the states indicated that they used holistic scoring procedures, with three states using primary trait techniques, one using analytical, and three states using both holistic and analytic scoring. In fact, when comparing their study with the earlier study by Fredrick, (1979), McCready and Melton found a change in preference from either holistic or primary trait scoring to a use of holistic and analytical methods, which appeared to offer a broader base for determining basic competencies in writing and assessment of educational progress.

The May 1984 issue of CAPTRENDS, published by the Center for Performance Assessment, reveals diverse environments for the solicitation of writing samples. Some states used untimed writing samples, while others set a 25-minute limit. Some states allowed

students to revise their initial drafts, while yet another state offered less skilled writers a number of prewriting suggestions to help them get started.

Large scale writing assessments are useful, but complex. This digest has attempted to identify a few of the issues and problems that need to be addressed in such an endeavor. However, as Spandel and Stiggins conclude in their booklet, Direct Measures of Writing Skills: Issues and Applications. Revised Edition (1981), "There is not now, nor will there ever be, a single best way to assess writing skills. Each individual educational assessment and writing circumstance presents unique problems to the developer and use of writing tests. Therefore, great care must be taken in selecting the approach and the methods to be used in each writing assessment. Methods used in one context to measure one state of relevant writing skill should not be generalized to other writing contexts without careful consideration of writing circumstances."

Holly O'Donnell, ERIC/RCS

Resources

Center for Performance Assessment. "Accurate Scoring, Quality Writing Prompts Improve State Practices," CAPTRENDS, 9(May 1984): 4-5.

Cronnell, Bruce, Ed.; Michael, Joan, Ed. "Writing: Policies, Problems and Possibilities." Proceedings of a Conference Co-sponsored by SWRL Educational Research and Development and by California State University, Los Alamitos, California, May 1982. ED 221 872.

Diederich, Paul B. Measuring Growth in English. Urbana, Ill.: National Council of Teachers of English, 1974. ED 097 702.

Fredrick, Vicki, "Writing Assessment Research Report: A National Survey." Madison, Wisc.: Wisconsin State Department of Public Instruction, 1979. ED 200 988.

Lloyd-Jones, Richard. "Primary Trait Scoring" in Evaluating Writing: Describing, Measuring, Judging, edited by Charles Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. ED 143 020.

McCaig, Roger A. "How to (and How Not to) Conduct an Assessment of Student Writing." 1982. ED 229 306.

McCready, Michael A.; Melton, Virginia S., "Feasibility of Assessing Writing Using Multiple Assessment Techniques. Research Report." Ruston, La.: Louisiana Technical University, 1981. ED 220 871.

Northwest Regional Educational Laboratory. "Writing Assessment for the 1980s. Proceedings of a National Conference on the Assessment of Writing." Boulder, Colorado, 1980. ED 195 576.

Quellmalz, Edys S. "Designing Writing Assessments: Balancing Fairness, Utility, and Cost." Los Angeles, Calif.: University of California, Center for the Study of Evaluation, 1982. ED 228 270.

Spandel, Vicki. Classroom Applications of Writing Assessment: A Teacher's Handbook. Portland, Ore.: Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, 1981. ED 215 995.

Spandel, Vicki; Stiggins, Richard J. Direct Measures of Writing Skill: Issues and Applications. Revised Edition. Portland, Ore.: Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, 1981. ED 213 035.

"Written Language and Writing Abilities: Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, July through December 1982 (Vol. 43, Nos. 1 through 6)." Urbana, Ill.: ERIC Clearinghouse on Reading and Communication Skills, 1982. ED 225 171.

ERIC

A Product of the ERIC Clearinghouse on Reading and Communication Skills
1111 Kenyon Road, Urbana, Illinois 61801
1984

This publication was prepared with funding from the National Institute of Education, U.S. Department of Education, under contract no. 400-83-0025. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to English language arts specialists for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the National Council of Teachers of English or the National Institute of Education.

ERIC Clearinghouse on Reading and Communication Skills

ERIC Digest

# Large Scale Writing Assessment

A national concern over the decline in students' writing scores (as revealed in National Assessment of Educational Progress reports), serious doubts about what some of those scores signify, and a shift in focus from writing product to writing process in research and classroom practice have each given impetus to the change from *indirect* measures of writing proficiency (those that use "objective" test items) to *direct* measures (those that call for student writing samples). In their 1981 national survey, McCready and Melton found that of the twenty-four states claiming to have a writing assessment program, twenty-two require a writing sample as part of the assessment. Only two states rely solely on the use of objective tests.

Large scale writing assessments, however, involve a number of complex issues that are not always evident to decision-makers who are not specialists in measurement. In discussing how to and how not to conduct an assessment of student writing, McCaig (1982) warns that "an assessment plan which is incomplete or poorly conceived may produce findings which can be challenged and even dismissed as meaningless by critics who can document flaws in the process." This digest (1) outlines some of the approaches used in the implementation of large scale writing assessments. (2) examines some of the issues and problems surrounding the use of student writing samples, and (3) reports on various trends in state writing assessment projects.

## Direct Versus Indirect Assessment

Direct and indirect writing assessments are radically different approaches focusing on different components of writing. Indirect measurements typically use multiple choice tests to assess the student's understanding of mechanics or language conventions: spelling, punctuation, capitalization, grammar, usage, sentence construction, organization, and so on. Direct assessments, on the other hand, assess actual writing performance, since they require the students to produce a writing sample. Spandel and Stiggins (1981) suggest that the two approaches can best be compared in terms of their advantages and disadvantages.

| | Advantages | Disadvantages |
|---|---|---|
| **Direct** | • extent of information provided about the students' writing proficiency | • potential lack of uniformity regarding proficiencies assessed |
| | • fidelity to real world writing tasks | • high cost of scoring |
| | • potential for positive user attitudes | |
| | • relatively low test development cost | |
| | • high face validity | |
| **Indirect** | • high score reliability | • lack of fidelity to real world writing tasks |
| | • relatively low scoring cost | • reliance on reading |
| | • high degree of control over skills tested | • lack of face validity |

Participants at a conference on assessment issues agreed that the use of writing samples is essential because of the instructional implications (McCready and Melton 1981). That is, if teachers know that students' writing ability will be evaluated by means of a direct measure, they will encourage more writing in the classroom.

## Three Approaches to Scoring

Provided that writing assessments are conducted to determine the status of writing in a given state or school system and provided they are conducted to help improve classroom instruction, several factors need to be considered prior to the collection of writing samples: (1) the educational decisions to be made, on the basis of test results; (2) the writing purpose, audience, and type of writing to be required; and (3) the specific skills or traits to be judged along with the criteria used for evaluating writing performance (Spandel and Stiggins 1981). It must also be remembered that ratings will vary depending upon the scoring procedure used. Quellmalz, in writing about scoring criteria (1982), notes that "criteria employed for evaluating student writing vary along a number of dimensions: from qualitative to quantitative; from general to specific; from comprehensive, full discourse features to isolated features; from vague guidelines to replicable, objective guidelines." Scoring options range from holistic scoring (general impressionistic marking) to analytic and primary trait scoring.

### Holistic Scoring

Holistic scoring of a writing sample is based upon the reader's overall impression of the effectiveness of a piece of writing. Papers are scored by trained raters on a numerical scale, usually a four-point scale. Once the writing samples are collected, the raters or scorers sort the samples into four stacks, relating the quality of the essay only to other papers in the group rather than to a predetermined example of "good" writing. Papers are typically read by two raters, and the scores they assign a writing sample are summed into a total score. If there is a discrepancy of two score points, the score is reconciled by yet a third reader/rater.

### Primary Trait Scoring

Primary trait scoring focuses on a specific rhetorical characteristic or trait of a given piece of writing. It is based on the premises that all writing is done in terms of a specific audience and that writing, if successful, will have the desired effect on that audience. Lloyd-Jones (1977) identifies two goals of primary trait scoring: (1) to define what segment of discourse will be evaluated (e.g., explanatory, expressive, persuasive), and (2) to train readers to render holistic judgments accordingly. A scoring guide for primary trait analysis may consist of the exercise itself; a description of the rhetorical traits of the writing; an interpretation of the exercise indicating how each element in the task is expected to affect the student; an interpretation of how the situation of the exercise is related to the primary trait; sample papers that are representative of each score point; and a discussion of why each sample paper was scored as it was (McCready and Melton 1981). One difference between holistic and primary trait scoring is that with primary trait scoring, students' papers are being measured against external criteria, whereas with holistic scoring, papers are compared with one another.

### Analytical Scoring

If primary trait scoring is a situation-specific analysis of writing, analytical scoring is a thorough, trait-by-trait analysis. The identified traits reflect those components of a writing sample that are considered important to any piece of writing in any context. Diederich (1974), the originator of analytical scoring, for example, has identified eight common traits: ideas, organization, wording, flavor (tone), usage, punctuation, spelling, and handwriting. Others may use traits more general such as content, organization, focus and support, and mechanics. If enough components are analyzed, this scoring procedure can provide a comprehensive picture of writing performance. However, the components need to be explicit and well defined so that the raters understand and agree upon the basis for making judgments about the writing sample.

In relating these scoring approaches to classroom applications, Spandel (1981) observes that holistic scoring offers a broad base for a discussion of what makes a piece of writing generally good or bad. Analytical scoring can take this discussion one step further by identifying those traits of components that make a piece of writing effective. And, by being situation-specific, primary trait scoring focuses on the importance of audience to a piece of writing.

### Issues and Problems

Essential to the quality of assessment and the value of scoring procedures used are the reliability and validity of the scores generated by the assessment. Specifically, the scoring criteria should be applicable uniformly within a rating session and

from one rating session to another. Furthermore, these ratings should correlate with other measures of student writing. Even if the assessment instrument is reliable and valid, spurious scores can result from the development of poor exercises, poor test administration or environment, or poor scoring procedures (Stiggins, Northwest Regional Educational Laboratory 1980). Scorers must understand and agree upon the criteria applied to a piece of writing. Thus, inadequate training of scorers may also influence or skew the results of an assessment.

The choice of topic (or "prompt") to be written is another factor that may influence scores. Students may write more enthusiastically on some topics than on others, resulting in better quality writing. A student's background and prior knowledge will also affect the "expertise" a student brings to a piece of writing. Or, depending upon the student's interpretation of a writing prompt, he or she may write persuasive discourse in response to a prompt intended for expository discourse.

Time and cost are two other factors that may influence the decision for large scale writing assessment. Stiggins (Northwest Regional Educational Laboratory 1980) separates cost factors into those that are developmental and those that are administrative. Developmental costs will vary depending on whether a previously designed assessment instrument is used or a new one developed. If one is to forego the expense of constructing a new scoring instrument, expense will still be incurred for the securing, reviewing, evaluating, and selecting of appropriate exercises and scoring guides that do exist. Administrative costs involve those associated with test administration, the selection of test administrators, the distribution of materials, and the collection of test materials.

Then there are the scoring costs—the time required to train raters and the time required to rate papers. According to data collected by Quellmalz (1982), the training time for holistic and primary trait scoring averages two to four hours, and for analytical scoring averages six to eight hours. Test reuse is another cost factor. Stiggins (Northwest Regional Educational Laboratory 1980) states that "with indirect assessment, the high cost of test construction can be amortized over repeated administration and the recurring scoring costs are very low. However, with direct assessment, although the initial development costs are low, the scoring costs remain high with each use." In a 1982 dissertation Bauer compared the reliabilities and the cost-efficiencies of these three methods of direct assessment. Based on her results, Bauer concluded that the analytical method was the most reliable and the holistic method was the most cost-efficient in grading a large number of essays (see ED 225 171).

A 1979 study by Fredrick identified some of the problems that states have encountered with their writing assessment endeavors: (1) arrangements for a place large enough and suitable enough for the scoring, (2) coordinating release time for teachers to act as scorers, (3) adhering to a rigid time schedule during the scoring session, (4) not enough time or money, and (5) finding or designing a variety of writing skill tasks. Her survey concluded with a list of recommendations to others who are planning a writing assessment, some of which are as follows:

- formulate writing objectives and focus research questions before the assessment,
- use language arts specialists to advise on content and to react to items prior to pilot testing,
- clarify traits to be measured,
- include clear and concise directions, and
- use actual performance on practical writing, such as messages, letters, forms and so forth, instead of the proofreading type of assessment found in most multiple choice tests.

## Trends in Writing Assessment

A national study conducted in 1981 by McCready and Melton collected data from 42 state departments of education. Of the 22 states using a writing sample, most of the states indicated that they used holistic scoring procedures, with three states using primary trait techniques, one using analytical, and three states using both holistics and analytic scoring. In fact, when comparing their study with the earlier study by Fredrick (1979), McCready and Melton found a change in preference from either holistic or primary trait scoring to a use of holistic and analytical methods, which appeared to offer a broader base for determining basic competencies in writing and assessing educational progress.

The May 1984 issue of CAPTRENDS, published by the Center for Performance Assessment, reveals diverse environments for the solicitation of writing samples. Some states used untimed writing samples, while others set a 25-minute limit. Some states allowed students to revise their initial drafts, while yet another state offered less skilled writers a number of prewriting suggestions to help them get started.

Large scale writing assessments are useful, but complex. This digest has attempted to identify a few of the issues and problems that need to be addressed in such an endeavor. However, as Spandel and Stiggins conclude in their booklet, *Direct Measures of Writing Skills: Issues and Applications. Revised Edition* (1981), "There is not now, nor will there ever be, a single best way to assess writing skills. Each individual educational assessment and writing circumstance presents unique problems to the developer and use of writing tests. Therefore, great care must be taken in selecting the approach and the methods to be used in each writing assessment. Methods used in one context to measure one state of relevant writing skill should not be generalized to other writing contexts without careful consideration of writing circumstances."

Holly O'Donnell, ERIC/RCS

## Resources


Center for Performance Assessment. "Accurate Scoring, Quality Writing Prompts Improve State Practices," *CAPTRENDS*, 9 (May 1984): 4–5.

Cronnell, Bruce, and Joan Michael, eds. "Writing: Policies, Problems and Possibilities." Proceedings of a Conference Co-sponsored by SWRL Educational Research and Development and by California State University, Los Alamitos, California, May 1982. ED 221 872.

Diederich, Paul B. *Measuring Growth in English*. Urbana, Ill.: National Council of Teachers of English, 1974. ED 097 702.

Fredrick, Vicki, "Writing Assessment Research Report: A National Survey." Madison: Wisconsin State Department of Public Instruction, 1979. ED 200 988.

Lloyd-Jones, Richard. "Primary Trait Scoring" in *Evaluating Writing: Describing, Measuring, Judging*, edited by Charles Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. ED 143 020.

McCaig, Roger A. "How to (and How Not to) Conduct an Assessment of Student Writing." 1982. ED 229 396.

McCready, Michael A., and Virginia S. Melton. "Feasibility of Assessing Writing Using Multiple Assessment Techniques. Research Report." Ruston: Louisiana Technical University, 1981. ED 220 871.

Northwest Regional Educational Laboratory. "Writing Assessment for the 1980s. Proceedings of a National Conference on the Assessment of Writing." Boulder, Colorado, 1980. ED 195 576.

Quellmalz, Edys S. "Designing Writing Assessments: Balancing Fairness, Utility, and Cost." Los Angeles: University of California, Center for the Study of Evaluation, 1982. ED 228 270.

Spandel, Vicki. *Classroom Applications of Writing Assessment: A Teacher's Handbook*. Portland, Ore.: Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, 1981. ED 215 995.

Spandel, Vicki, and Richard J. Stiggins. *Direct Measures of Writing Skill: Issues and Applications. Revised Edition*. Portland, Ore.: Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, 1981. ED 213 035.

"Written Language and Writing Abilities: Abstracts of Doctoral Dissertations Published in *Dissertation Abstracts International*, July through December 1982 (Vol. 43, Nos. 1 through 6)." Urbana, Ill.: ERIC Clearinghouse on Reading and Communication Skills, 1982. ED 225 171.



ERIC

A Product of the ERIC Clearinghouse on Reading and Communication Skills
1111 Kenyon Road, Urbana, Illinois 61801
1984

This publication was prepared with funding from the National Institute of Education, U.S. Department of Education, under contract no. 400-83-0025. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to English language arts specialists for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the National Council of Teachers of English or the National Institute of Education.